Doctoral Thesis

Refining the genetic architecture of Atlantic salmon (*Salmo salar*) maturation using genomics-enabled approaches

Marion Sinclair-Waters

Doctoral Programme in Wildlife Biology
Organismal and Evolutionary Biology Research Programme
Faculty of Biological and Environmental Sciences
University of Helsinki

DOCTORAL DISSERTATION

To be presented for public examination with the permission of the Faculty of Biological and Environmental Sciences of the University of Helsinki, in lecture room 2041, Biocentre 2, on November 5th, 2021, at 13:15.

Helsinki 2021

The Faculty of Biological and Environmental Sciences uses the Urkund system (plagiarism recognition) to examine all doctoral dissertations.

ISBN 978-951-51-7525-0 (paperback)
ISBN 978-951-51-7526-7 (PDF)

Unigrafia
Helsinki 2021

Supervisors

Professor Craig Primmer
Organismal and Evolutionary Biology Research Programme
Institute of Biotechnology
University of Helsinki

Dr. Nicola Barson
Centre for Integrative Genetics
Department of Animal and Aquaculture Sciences
Norwegian University of Life Sciences

Thesis advisory committee

Dr. Philine Feulner
Department of Fish Ecology and Evolution
Eawag - Swiss Federal Institute of Aquatic Science and Technology

Dr. Tanja Pyhäjärvi
Department of Forest Sciences
Faculty of Agricultural and Forestry
University of Helsinki

Reviewed by

Professor Jon Slate
Department of Animal & Plant Sciences
University of Sheffield

Dr. Nina Overgaard Therkildsen
Department of Natural Resources and the Environment
Cornell University

Opponent

Professor Catherine Peichel
Institute of Ecology and Evolution
University of Bern

Custos

Professor Craig Primmer
Organismal and Evolutionary Biology Research Programme
Institute of Biotechnology
University of Helsinki

Table of Contents

1. Abstract

Understanding the genetic architecture of adaptive traits is a common goal in evolutionary biology. However, there are relatively few well-characterized genetic architectures for adaptive traits in non-model systems, particularly in wild populations. In this thesis, we further characterize the genetic architecture of age at maturity in Atlantic salmon using a combination of genomic-enabled approaches. First, a large-scale genome-wide association study of ~11000 male Atlantic salmon, with ~500K single nucleotide polymorphisms (SNPs) genotypes, was used to identify loci underlying age at maturity. Significant associations were found on 28 of the 29 Atlantic salmon chromosomes, including two strong signals at the *six6* and *vgll3* gene regions. Furthermore, 116 candidate loci with varying effect sizes were identified revealing a mixed genetic architecture with a combination of large-effect loci and a polygenic component consisting of multiple smaller-effect loci. A combination of multi-SNP association methods and individual-level sequencing data was then used to resolve the genetic architecture within the 116 candidate loci for Atlantic salmon age at maturity. The genetic architecture within these candidate loci was relatively simple, where the association was driven by a single mutation. Finally, ~6000 male Atlantic salmon with a high frequency of recombinant alleles at a known large-effect locus spanning two linked genes, *vgll3* and *akap11*, were reared to maturation age. *Vgll3* SNP variation was strongly associated with maturation timing, however, *akap11* SNP variation showed little to no association with maturation timing, suggesting that variation linked to *vgll3* is driving the association at this large-effect locus. Together, these findings further refine our understanding of the genetic architecture for Atlantic salmon age at maturity. Additionally, this thesis provides an analytical framework for characterizing the genetic architecture of adaptive traits in non-model systems.

2. Tiivistelmä

Tässä väitöskirjassa hyödynnetään genomisia lähestymistapoja Atlantin lohen sukukypsyysiän geneettisen arkkitehtuurin selvittämisessä. Lisäksi väitöskirjatyöni tarjoaa analyyttisen kehyksen eri piirteiden geneettisen arkkitehtuurin luonnehtimiseksi muissa kuin mallilajeissa. Ensimmäisessä työssä laajamittainen genomin laajuinen assosiaatiotutkimus paljasti Atlantin lohen sukukypsyysiän geneettisen arkkitehtuurin muodostuvan yhdistelmästä suuri- ja pienivaikutteisia lokuksia. Toisessa työssä paljastimme assosiaatiomenetelmiä ja perimäsekvensointidataa hyödyntämällä ehdokaslokusten sisäisen arkkitehtuurin melko yksinkertaiseksi ja assosiaation selittyvän yhdellä mutaatiolla. Kolmannessa työssä erotimme kytkeytyneiden *vgll3*- ja *akap11*-geenien vaikutukset kasvattamalla jälkeläisiä, joilla oli eri rekombinanttialleeliyhdistelmät kyseisessä kromosomissa 25 sijaitsevassa suurivaikutteisessa lokuksessa. Työ paljasti, että *akap11*-vaihtelulla ei ole juurikaan vaikutusta sukukypsyysikään, ja että assosiaatio on seurausta muuntelusta lähempänä *vgll3*-geeniä.

3. List of Original Publications

This thesis is based on the following publications, which are referred to in the text by their Roman numerals:

I **Sinclair-Waters, M**., Ødegård, J., Korsvoll, S.A., Moen T., Lien, S., Primmer, C.R., & Barson, N.J. Beyond large-effect loci: Large-scale GWAS reveals a mixed large-effect and polygenic architecture for age at maturity of Atlantic salmon. *Genetics, Selection, Evolution*. (2020)

II **Sinclair-Waters, M.**, Nome, T., Wang, J., Lien S., Kent, M.P., Sægrov, H., Florø-Larsen, B., Bolstad, G.H., Primmer, C.R* & Nicola J. Barson, N.J.* Dissecting the loci underlying maturation timing in Atlantic salmon using haplotype and multi-SNP based association methods. *BioRxiv.* (2021)

III **Sinclair-Waters, M.**, Piavchenko, N., Ruokolainen, A., Aykanat, T., Erkinaro, J. & Primmer, C.R. Refining the genomic location of SNP variation affecting Atlantic salmon maturation timing at a key large-effect locus. *BioRxiv.* (2021)

*Authors contributed equally

4. Author Contributions

I. NJB, TM, JØ conceived the idea for the study. All authors contributed to aspects of the study design and the completion of the work. SAK designed and conducted the fish phenotyping. MSW, NJB, SL developed the analysis methods and performed analyses. TM and JØ contributed phenotype and genotype data for the study. MSW, CRP and NJB wrote the manuscript.

II. CRP, NJB, MSW conceived the study. TN developed the variant calling workflow and constructed the fixed assembly of ssa25. JW developed the variant filtering criteria. MSW performed all downstream analyses with input from NJB. MPK played key role in generating whole genome sequencing data. SL led the whole genome sequencing work as part of the AquaGenome project. HS, GB, BFL, CRP coordinated Atlantic salmon sampling and provided phenotypic information. MSW, CRP, NJB wrote the manuscript.

III. CRP, MSW conceived the study. MSW, CRP designed crosses. CRP, NP, MSW designed experimental setup. NP supervised fish husbandry and maintenance of fish-raising facility. MSW, NP led tagging, tissue collection, and phenotypic data collection. AR, TA, CRP developed the KASP genotyping protocol. AR performed genotyping. MSW performed genotype calling and data analysis. JE provided parental material for crosses. MSW, CRP wrote the manuscript.

5. Abbreviations

GRM	genomic relatedness matrix
GWAS	genome-wide association study
LD	linkage disequilibrium
SNP	single nucleotide polymorphism
PVE	proportion of variance explained

6. Introduction

A fundamental goal in evolutionary biology is to better understand the genetic underpinnings of phenotypic variation. Such an understanding can improve our ability to predict how organisms will respond to environmental changes and disturbances. Additionally, a well-characterized genotype-phenotype link helps to understand how phenotypic diversity is maintained in a population and is therefore important for designing conservation and management strategies aimed at maintaining such diversity. Despite its importance, the link between genotype and phenotype remains unknown for many traits, particularly for traits observed in wild populations. Given our limited understanding of the important relationship between phenotype and genotype, many critical questions remain: How many loci are involved? Where in the genome are these loci located? What are the effect sizes of these loci? And is the locus effect driven by a single or multiple mutation(s)? The recent growth in available genomic data has created new opportunities for exploring the genotype-phenotype relationship.

6.1 Genetic architecture

Genetic architecture is a broad term used to describe the collection of genetic features that are involved in controlling a trait or fitness. The genetic architecture of a trait is a description of the number of loci, the genomic location of these loci, their effect sizes and the allele frequencies at loci underlying trait variation. In this thesis, we will focus on these key features of genetic architecture. Genetic architecture, however, also includes the mode of gene action (e.g. additive, dominant or recessive), the epistatic interactions between or among loci, linkage among loci, and how loci effects depend on environmental conditions (i.e. gene by environment interactions).

Genetic architecture can be described as monogenic, oligogenic, or polygenic, whereby the trait is controlled by one, a few, or many loci, respectively. For traits such as pigmentation (Hoekstra 2006; Linnen *et al.* 2013), lactase persistence (Tishkoff *et al.* 2007), and insecticide resistance (Daborn *et al.* 2002), genetic architecture is

relatively simple, where the trait is controlled by a single or small number of large-effect loci. In contrast, many traits, including many human diseases (Purcell *et al.* 2009; Loh, Bhatia, *et al.* 2015) and height (Pritchard and Di Rienzo 2010), have a polygenic architecture. Further, the omnigenic model builds upon the polygenic model and proposes that a large fraction of all genes can affect a trait via a vast network of core genes that directly affect the trait and many peripheral genes that mediate core gene effects (Boyle, Li, and Pritchard 2017; Liu, Li, and Pritchard 2019). Characterizing genetic architecture is crucial for determining how trait variation may be altered via processes such as selection, genetic drift and/or gene flow (Kuparinen and Hutchings 2017; Oomen, Kuparinen, and Hutchings 2020; Le Corre and Kremer 2012). Additionally, the genetic architecture of a trait can provide insight into how genetic and trait variation are maintained in a population or species (Barson *et al.* 2015; Turelli and Barton 2004).

6.1.1 Genetic architecture of adaptive traits

For traits under selection, theoretical models suggest that favourable alleles will cluster together forming tightly linked loci, with relatively large effect sizes. This clustered genetic architecture is expected to arise in spatially and/or temporally heterogeneous environments where populations are experiencing a specific balance of gene flow and selection (Yeaman and Whitlock 2011; Yeaman 2013; Yeaman, Aeschbacher, and Bürger 2016). Indeed, several studies on wild populations have reported strong signals of selection and/or association. These include a large-effect locus for sea age at maturity in Atlantic salmon (*Salmo salar*) (Barson *et al.* 2015) and a structural variant strongly associated with mating strategies in ruffs (*Philomachus pugmax*) (Lamichhaney et al. 2015), among others (Gratten et al. 2007; Joron et al. 2006; Lamichhaney et al. 2017). It is plausible that an interplay between migration and selection has contributed to the establishment of these major loci. Additional modelling and empirical studies on a wider variety of systems will help to elucidate the evolutionary processes leading to genetic architectures consisting of large-effect loci.

The genetic architecture of a trait is highly relevant when predicting how trait variability in a population will change in response to environmental changes and other selective pressures. During environmental change causing a shift in a phenotypic

optimum, population viability is higher when genetic architecture is polygenic rather than monogenic/oligogenic (Kardos and Luikart 2021). This may be due to heritability and rate of adaptation declining more rapidly for traits with large-effect loci than polygenic traits (Barton and Keightley 2002). Under directional selection pressure, trait variation decreases when the genetic architecture is polygenic. In contrast, the outcome for a trait controlled by single-locus is highly variable (Oomen, Kuparinen, and Hutchings 2020; Kuparinen and Hutchings 2017). Furthermore, traits with polygenic architectures, where causal loci are highly linked due to recombination suppression (e.g. genomic rearrangement), respond similarly to traits with single large-effect locus architectures under periods of directional selection (Oomen, Kuparinen, and Hutchings 2020). Ultimately, however, the ability to model such responses to selection depends largely on a well-characterized genetic architecture that includes a thorough understanding of the number of loci controlling the trait, their location in the genome, and their effect sizes.

6.1.2 Characterizing genetic architecture

A common first step to characterize genetic architecture is to do a genome-wide association study (GWAS) to identify genetic variation, typically single nucleotide polymorphisms (SNPs), that is associated with trait variation. Methods used for association testing can test each SNP independently (single-SNP approaches) or can consider the effect of other SNPs when testing a SNP for an association (multi-SNP approaches). Given the causal SNP or a SNP in linkage disequilibrium (LD) with the causal SNP is represented in the data, large-effect loci with strong trait associations are relatively easy to detect. Small-effect loci or rare alleles, however, can be challenging or even impossible to detect. Identification of such loci may require higher detection power, which can be reached by increasing the sample size of individuals, but this level of sampling may not be feasible (Sham and Purcell 2014).

Over the past few decades, an increase in sequencing technologies and bioinformatics tools, as well as a decrease in their costs, have made using GWAS to characterize the genetic architecture of traits more feasible, particularly in non-model systems. Furthermore, methodological developments and tools related to GWAS are increasingly available, enabling the analysis of traits with a variety of distributions, helping to account for population structure and relatedness among samples, and

decreasing computational costs (Zhou *et al.* 2018; Loh, Tucker, *et al.* 2015; Euahsunthornwattana *et al.* 2014; Legarra, Ricard, and Varona 2018; Chen *et al.* 2016; Bulik-Sullivan *et al.* 2015; Listgarten *et al.* 2012; Price *et al.* 2010). Additionally, methods to test the cumulative effects of multiple SNPs, rather than a single SNP, can be informative when a single-SNP's association is weak and does not pass the significance threshold, but when considered in combination with other SNPs has a significant association with the trait. Multi-SNP approaches can also help to resolve how many mutations underlie an association signal. Such developments in methods for association studies, particularly when applied to wild populations, provide an opportunity for significant progress in characterizing genetic architecture.

Following the detection of an association via GWAS, a common next step is to identify candidate gene(s) in the associated region. Often the associated region contains multiple genes and therefore it is unclear which gene or genes is/are linked to the association signal. In such cases, validation of candidate genes is required. In model systems, validation of candidate genes can be done using gene knockouts or genome editing technologies (e.g. CRISPR) (Varshney *et al.* 2013; The International Mouse Knockout Consortium 2007; Sander and Joung 2014). Candidate gene validation via CRISPR has been successful in some free-living species (Livraghi *et al.* 2018; Concha *et al.* 2019; Woronik *et al.* 2019; Wucherpfennig, Miller, and Kingsley 2019; Rodríguez-Leal *et al.* 2017; Sedeek, Mahas, and Mahfouz 2019), however, remains particularly challenging or unachievable in many free-living species. Alternatively mapping approaches, which take advantage of naturally occurring recombination events in the associated region, can be used for delineating which gene(s) is/are linked to association signals. This approach, however, is also challenging as it requires screening of many individuals in order to identify informative recombination events, followed by controlled crossing of such individuals and phenotypic assessment.

6.2 Atlantic salmon (*Salmo salar*) age at maturity as a study trait

6.2.1 Atlantic salmon age at maturity

Atlantic salmon is an anadromous species that exhibits a diversity of life history-strategies. Variation in age at maturity contributes substantially to this diversity. Atlantic salmon can spend anywhere from one to seven years in freshwater prior to migrating to the marine environment. Once at sea, individuals can spend anywhere from one to five years before reaching maturation and returning to their natal rivers to spawn. Furthermore, some individuals (mostly males), known as mature parr, will mature in freshwater without having migrated to sea (Mobley et al. 2021; Erkinaro et al. 2019). Maturation at a later age is associated with larger body size that increases fecundity of females and reproductive success of males. Late maturation, however, increases the risk of mortality before reproduction. In contrast, early maturing individuals are smaller and have less reproductive success at spawning grounds, but have a higher chance of surviving to reproductive age (Mobley et al. 2020; Fleming and Einum 2011).

In aquaculture salmon, variation in age at maturity is also observed. In fact, there is considerable interest in understanding the genetic basis of age at maturity in aquaculture salmon due to the negative effects of early maturation. Early maturation usually refers to maturation occurring 1-year post-smoltification (where smoltification refers to the process inducing morphological, physiological, and behavioural changes needed for the transition from freshwater to the marine environment) (Mobley et al. 2021). Early maturation results in lost revenue (McClure *et al.* 2007) due to flesh degradation that occurs during the maturation process (Aksnes, Gjerde, and Roald 1986), and can negatively affect the health of aquaculture fish (Taranger *et al.* 2010).

6.2.2 Genetic basis of Atlantic salmon age at maturity

Given that age at maturation in Atlantic salmon is highly heritable (Gjerde 1984; Reed *et al.* 2015), there has been considerable interest in identifying the genes underlying maturation in Atlantic salmon. A combination of quantitative trait loci (QTL) mapping and GWAS has improved our understanding of the genetic basis over the last couple of decades (Mobley et al. 2021). For example, a QTL study

identified loci on chromosomes 10 and 21 for early male maturation (Gutierrez *et al.* 2014), however did not have sufficient genomic coverage to identify potential genes involved. Some of the first GWAS, using ~6K SNPs, found associations in both wild and aquaculture strains identified loci on a number of chromosomes in both wild and aquaculture strains (e.g. 1, 2, 9, 10, 12, 13, 16, 21, 25, 27, 28). However, only ~500 individuals and ~4K SNPs were used and these studies therefore lacked statistical power and genomic coverage to provide a clear picture of the genomic regions underlying age at maturity (Johnston *et al.* 2014; Gutierrez *et al.* 2015).

Following these earlier mapping studies, GWAS with larger sample sizes and genomic coverage identified a large-effect locus on chromosome 25 that explains up to 39% of the variation in sea age at maturity (Ayllon *et al.* 2015; Barson *et al.* 2015). The GWA study by Barson *et al.* (2015), consisting of 1404 individuals sampled from 57 European populations, detected the association in the Atlantic, Barents/White and Baltic Sea lineages, showing that the locus effect is conserved across all three European phylogeographic lineages. The primary candidate gene located upstream of the top-associated SNP on chromosome 25 is *vgll3* (vestigial-like family member 3). The *vgll3* gene encodes a transcription cofactor involved in regulating adipogenesis (Halperin *et al.* 2013) and is associated with human maturation (Perry *et al.* 2014; Day *et al.* 2017; Cousminer *et al.* 2013), suggesting it is plausible candidate for being involved in maturation in Atlantic salmon. A second gene located downstream of the top-associated SNP, *akap11* (A-kinase anchoring protein 11), is also a plausible candidate gene given it is expressed in testes during spermatogenesis and affects sperm motility in mice and humans (Reinton *et al.* 2000; Luconi *et al.* 2004; Miki *et al.* 2002). In North American populations, a higher proportion of late maturation alleles of *vgll3* are found in late-maturing females than in early-maturing individuals, providing some evidence that *vgll3* may also play a role in maturation in this lineage (Kusche *et al.*, 2017). Association studies for maturation in aquaculture strains have found an association with chromosome 25 in some strains (Gutierrez *et al.* 2015; Ayllon *et al.* 2015), but not others (Boulding *et al.* 2019). Additionally, in some cases the association is sex-specific where it is observed only in males (Ayllon *et al.* 2019) or only in females (Mohamed *et al.* 2019). Such differences may be due to different genetic architectures of the trait among populations/strains. Alternatively, this may be due to low genetic variation at

the *vgll3* locus in these populations/strains and/or insufficient power to detect such an association. To better understand the genetic architecture of this trait, we require higher-powered association studies involving a larger number of individuals and a broad range of populations and aquaculture strains.

Barson *et al.* (2015) identified another candidate gene on chromosome 9, *six6*, that was strongly associated with maturation prior to population stratification correction, after which the association signal significance was lost. The lost signal may be because the locus reflects well population structure, possibly due to being involved in local adaptation (Bourret *et al.* 2013; Moore *et al.* 2014; Pritchard *et al.* 2018), or that the locus' effect varies among populations. To resolve the genotype-phenotype association of loci involved in local adaptation, association studies within a single population may be required. The primary candidate gene, *six6*, in is a transcription factor of the hypothalamus-pituitary-gonadal axis and is also associated with height and age at maturity in humans (Day *et al.*, 2017; Perry *et al.*, 2014) and in cattle (Cánovas *et al.*, 2014).

There has been substantial improvement in our understanding of the genetic basis of age at maturity in Atlantic salmon, which has been largely due to advancement in tools allowing for the genotyping of a greater number of markers in a larger sample of phenotyped individuals. Nevertheless, there remains some uncertainty concerning the genetic architecture of age at maturity in Atlantic salmon.

7. Aims of the Thesis

The overall aim of this thesis is to further refine the genetic architecture of Atlantic salmon age at maturity. The work is organized around three distinct aims:

1) To characterize the genetic architecture of Atlantic salmon age at maturity using a large-scale genome-wide association study (GWAS).
2) To use haplotype-based and multi-SNP association methods to further characterize known candidate regions for age at maturity in Atlantic salmon.
3) To more precisely map causal SNP variation at a large-effect locus for age at maturity in Atlantic salmon.

8. Materials and Methods

8.1 Animal material and maturation phenotyping

For **Chapter I**, 11,166 male Atlantic salmon were sampled from the Norwegian AquaGen aquaculture strain. This strain is a result of a breeding program that began in the 1970's with founder individuals originating from 41 Norwegian rivers (Gjedrem, Gjøen, and Gjerde 1991). Maturation phenotype was determined based on an individual's maturity status one year after transfer to sea and was scored as a binary trait: mature (early maturing individual) or non-mature. Maturity status was determined by checking for the presence of a developed kype and darkened colouration, which are characteristics of maturation.

For **Chapter II**, 313 Atlantic salmon were collected from 53 Norwegian and Finnish rivers. Maturation phenotype was measured as the number of years spent at sea prior to maturation and return to river for spawning (sea age at maturity). Individuals were scored based on sea age determined using scale readings (ICES 2011): 1 (one year spent at sea), 2 (two years spent at sea), or 3 (three or more years spent at sea). Only five individuals had a sea age greater than three and were therefore combined with three-year individuals for analyses.

For **Chapter III**, controlled crosses using parental material from a captive broodstock, originating from the Neva river population, were used to breed progeny with a high number of recombinant alleles at a large-effect locus on chromosome 25 (Barson *et al.* 2015). In total, 5895 male Atlantic salmon were reared to one year of age. In captivity, male Atlantic salmon are able to mature at this age (Debes *et al.* 2019). Maturation status was scored as a binary trait, mature or non-mature, based on internal examination of gonad size and colour. White and enlarged gonads (filling at least 75% of the body cavity) were considered signs of maturation.

8.2 Genetic markers

Individuals used in **Chapter I** were genotyped using a 50K SNP array and imputed up to 512,397 SNPs using parents genotyped with a 930K SNP array and pedigree information. **Chapter II** used individual full-genome sequences described in

Bertolotti *et al.* (2020). SNP genotypes were called using GATK according to GATK Best Practices workflow (Depristo *et al.* 2011). Parental material used in **Chapter III** was genotyped at 177 SNPs described in Aykanat *et al.* (2016). Progeny were genotyped at three loci using the KASP genotyping method (He, Holme, and Anthony 2014): top-associated *vgll3* SNP, *akap11* missense SNP (Barson *et al.* 2015), and *SDY* locus (Yano *et al.* 2012).

8.3 Association testing

Genome-wide association testing in **Chapter I** used a linear mixed model method implemented in BOLT-LMM (Loh, Tucker, *et al.* 2015). The model included a genomic relatedness matrix (GRM) as a random effect to account for relatedness and population structure in the samples. Only SNPs showing no evidence LD and not on the same chromosome as the SNP being tested were used for GRM calculation to avoid redundant genotypic information and proximal contamination, respectively.

Many SNPs in an associated region display a significant association due to LD with the causal SNP. To determine which SNPs represented independent signals from the BOLT-LMM analysis in **Chapter I**, we used conditional and joint analysis implemented in GCTA (Yang *et al.* 2011). This approach adjusts for the effects of neighbouring SNPs when testing for an association (Yang *et al.* 2012). Additionally, variance component partitioning, implemented in BOLT-REML (Loh, Tucker, *et al.* 2015), was then used to determine the proportion of variance explained by resulting set of significant SNPs identified using conditional and joint analysis.

In **Chapter II**, association testing of SNPs within 500kb regions surrounding the candidate loci identified in **Chapter I** was used to further resolve the SNPs underlying the association of these regions with sequence-level data. Two multi-SNP association methods were used: haplotype association testing implemented in *hapQTL* (Xu and Guan 2014; Guan 2014) and Bayesian Variable Selection regression implemented in *PiMASS* (Guan and Stephens 2011). Single-SNP association testing with *hapQTL* was also used.

In **Chapter III**, associations of *vgll3* and *akap11* SNP variation were tested using mixed-effect logistic regression implemented in the *lme4 R* package (Bates *et al.* 2015).

8.4 Identifying candidate genes

The closest gene (and within 50kb) to a significantly associated SNP following condition and joint analysis in **Chapter I** was considered a candidate gene for age at maturity.

9. Results and Discussion

9.1 Age at maturity has a mixed large-effect and polygenic architecture

Genome-wide association testing in **Chapter I** identified 13,149 SNPs that were significantly associated with maturation timing. The effect sizes of SNPs ranged from large to small or no effect. Some association peaks were broad due to slow LD decay in regions of the genome, which is characteristic of aquaculture Atlantic salmon strains that have smaller effective population sizes relative to wild populations (Hindar et al. 2006; Waples, Larson, and Waples 2016). Condition and joint analysis identified 116 SNPs that were independently associated with the trait (Figure 1).

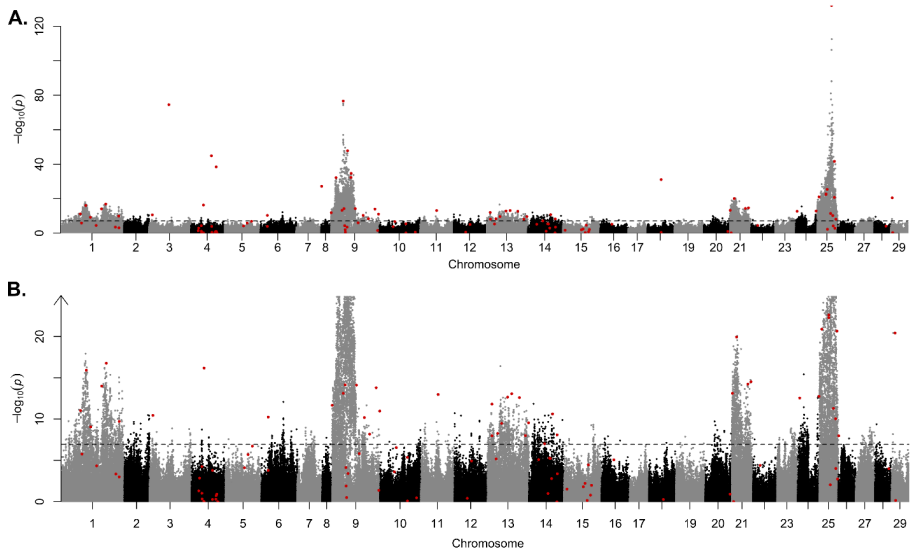

Figure 1. Manhattan plots of GWAS for age at maturity in males using 11,116 individuals. Red dots mark loci that were significant after conditional and joint analysis. Significance threshold was adjusted to account for multiple testing using Bonferroni correction. A. Manhattan plot showing all SNPs and their corresponding association statistics. B. Zoomed view of SNPs with association statistics below a $-\log_{10}(P\text{-value})$ of 25. Y-axis is truncated at $-\log_{10}(P\text{-value})$ 25.

The proportion of variance explained by the set of 116 independently associated SNPs was determined using variance component partitioning. The 116 SNPs

explained 78% of the genetic variance and the other ~500K SNPs explained the remaining 22%. This suggests our GWAS framework was adequately-powered to identify the genomic location of a substantial proportion of variation underlying the trait. It is unknown, however, how many of the remaining ~500K SNPs have an effect on maturation. It is possible that many additional causal SNPs exist with very small effects on maturation.

A highly significant SNP in the *vgll3* region on chromosome 25 confirmed the importance of a large-effect locus previously identified in other studies (Barson *et al.* 2015; Ayllon *et al.* 2015). Another highly significant SNP was located on chromosome 9 in close proximity to *six6*. This locus was also identified in Barson *et al.* (2015), however only prior to correction for population stratification. This highlights the value of single population association studies to identify trait-associated loci that also play a role in local adaptation.

The findings presented in **Chapter I** provide evidence that Atlantic salmon age at maturity has a mixed genetic architecture with a few larger effect loci and a polygenic background consisting of multiple smaller-effect loci. The characterization of the genetic architecture of this trait contributes to our understanding of how Atlantic salmon can respond to selective pressures. Furthermore, this mixed architecture broadens the scope of empirical examples for genetic architectures. Such mixed-architectures are challenging to characterize given the difficulty of detecting small-effect loci, particularly in wild populations. Consequently, it is possible that this type of architecture is more common than what is reported and should therefore be considered in analytical frameworks for characterizing the genetic architecture of traits in wild populations.

9.2 Multi-SNP analyses resolve number and identity of associated SNPs at candidate loci

Single-SNP and haplotype association analyses in **Chapter II** found strong association signals (Bayes factor > 3) with sea age at maturity at 5 of the 116 candidate regions from **Chapter I** (Table 1). The lack of an association within the other regions may be due to differences in genetic architecture between wild and aquaculture Atlantic salmon. Alternatively, such differences may be due to the **Chapter II** study

having lower detection power than the **Chapter I** study, particularly for the detection of loci with small effect sizes and/or rare alleles. In all five cases where a strong association signal was found, the strongest association was with a single SNP rather than a haplotype (Table 1). This suggests the association may have arisen via a single mutation, opposed to multiple mutations or recombination introducing new combinations of existing mutations.

Table 1. Strongest association signals from *hapQTL* for each candidate region showing evidence of an association with sea age at maturity and the closest gene. Top SNPs and candidate gene from **Chapter I** are also listed.

Candidate region	**Top signal**	**Closest gene**	**Top SNP(s) (Chapter I)**[a]	**Candidate gene(s) (Chapter I)**
ssa06:27541960-28218141	6:28045390	*pecam1*	6:27791960 6:27968141	*slc9a3r1* *recql5* LOC106606978
ssa09:10915066-11415066	9:11266848	*asap2a*	9:11165066	*mboat2*
ssa09:24636574-25136574	9:24888841	*six6*	9:24886574	*six6*
ssa21:49390687-49890687	21:49645222	*taar13c*	21:49640687	*taar13c*
ssa25:28389273-28889273	25:28651640 [ICSASG_v2: 25:28669350]	*vgll3*	25:28910202	*vgll3*

Bayesian Variable Selection regression implemented in *PiMASS* was used to further analyse the five candidate regions showing a strong association in the *hapQTL* analyses. The *PiMASS* analyses found that the number of SNPs contributing to variance in sea age at maturity was one for all candidate regions tested, excluding the ssa09:24636574-25136574 region, where the number of SNPs was two. After regressing out the top-associated SNP from the phenotype values and repeating the analysis, however, the mode number of SNPs was zero for all candidate regions. Furthermore, the median proportion of variance explained was 0% or 1% for all candidate regions (Table 2). This suggests that variation in sea age at maturity explained by each of these five candidate region is largely driven by a single SNP.

Table 2. *PiMASS* results prior to and after regression of top-associated SNP identified in the initial *PiMASS* analysis. These include the mode of the distribution of the number of SNPs and the median of the distribution of proportion of variance explained (PVE) for a model explaining sea age at maturity.

Candidate region	Mode # of SNPs	Median PVE	Mode # of SNPs (post-regression)	Median PVE (post-regression)
ssa06:27541960-28218141	1	0.05	0	0
ssa09:10915066-11415066	1	0.07	0	0.01
ssa09:24636574-25136574	2	0.09	0	0.01
ssa21:49390687-49890687	1	0.04	0	0
ssa25:28389273-28889273	1	0.19	0	0.01

Our findings provide evidence that loci underlying sea age at maturity in Atlantic salmon can be predominantly driven by a single SNP and thus have relatively simple and non-clustered alleles. Contrastingly, in other species, such as *Peromyscus maniculatus* (Linnen *et al.* 2013), *Drosophila melanogaster* (Bickel, Kopp, and Nuzhdin 2011), *Arabidopsis thaliana* (Kerdaffrec *et al.* 2016), and *Gasterosteus aculeatus* (Archambeault *et al.* 2020), complex genetic architectures within trait-associated loci have been characterized, where multiple clustered mutations have independent effects on the trait or fitness. Additionally, theoretical modelling predicts such clustered loci under certain levels of gene flow and selection. The different genetic architecture within age at maturity loci may be due to gene flow among Atlantic salmon populations being too restricted and/or the strength of selection being insufficient for the establishment of clustered loci. Additional modelling and more empirical data on a variety of traits and species may help to elucidate when and why certain genetic architectures arise.

9.3 Natural recombination enables more precise mapping of causal variation at large-effect locus

A total of 5895 males were reared to maturation age. Of these 5895 individuals, 4769 had recombinant genotypes (i.e. carrying a haplotype with an *L* allele for vgll3 and an *E* allele for akap11, or vice versa). The overall maturation rate was 2.87%. The *vgll3 EE* and *EL* genotypes increased the *log*(odds ratio) of maturation relative to the *LL* genotype by 4.21 and 1.79, respectively. For *akap11*, only the *EE* genotype had a significant effect, which was negative and decreased the *log*(odds ratio) of maturation by 1.30, relative to the *LL* genotype (Figure 2).

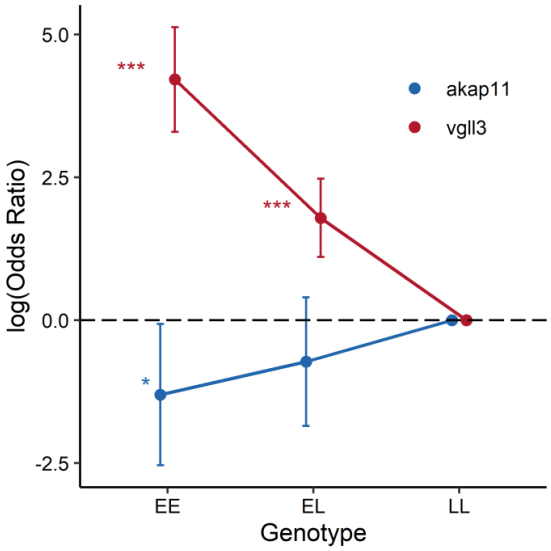

Figure 2. Ratio of the odds of maturation on the logarithmic scale and the respective 95% confidence intervals of the *EE* and *EL* genotypes for *vgll3* and *akap11*, relative to the *LL* genotype. Asterisks denote level of significance (* *p*-value < 0.05, *** *p*-value < 0.001).

Chapter III shows that the association at a key large-effect locus for age at maturity on chromosome 25 is explained by *vgll3* SNP variation, and not *akap11* SNP variation. This narrows down the genomic location of causal SNP variation for Atlantic salmon age at maturity and provides stronger evidence that *vgll3* is the primary candidate gene at the chromosome 25 locus. These findings will help guide future research determining how this locus is involved in genetic processes affecting Atlantic salmon maturation. For example, *vgll3* would be an appropriate target for

CRISPR genome editing to further resolve the effect of this gene on Atlantic salmon maturation. Furthermore, **Chapter III** highlights the utility of natural recombinants to more precisely map causal SNP variation, particularly when genome editing is not feasible.

10. Conclusions

Throughout this thesis, I have used genetic and genomic approaches to characterize different aspects of the genetic architecture of Atlantic salmon age at maturity. Additionally, this thesis provides an analytical framework for characterizing the genetic architecture of traits in non-model systems.

There are three main findings of this thesis. First, a large-scale genome-wide association study revealed a mixed genetic architecture of age at maturity in Atlantic salmon with a combination of large-effect loci and a polygenic background consisting of smaller-effect loci. Second, a combination of multi-SNP association methods and individual-level sequencing data revealed relatively simple, non-clustered, genetic architecture within candidate loci, where the association was driven by a single mutation. Finally, we delineate the effects of two linked genes, *vgll3* and *akap11*, at a large-effect locus on chromosome 25 by rearing progeny with recombinant alleles at this locus. This revealed that *akap11* SNP variation had little to no effect on age at maturity and that the association of this key locus is a result of SNP variation in closer proximity to *vgll3*. These three findings further refine our understanding of the genetic architecture for Atlantic salmon age at maturity.

11. Acknowledgements

There are many people I must thank for helping me with my PhD research.

First and foremost, I thank my supervisors Prof. Craig Primmer and Dr. Nicola Barson for their guidance and feedback throughout this project. I have learned a tremendous amount from you both. Also, thank you for the many opportunities to grow as a researcher.

I thank the members of my advisory committee, Dr. Philine Feulner and Dr. Tanja Pyhäjärvi, for making sure the project was moving along smoothly and your kind input. I extend my thanks to Prof. Jon Slate and Dr. Nina Therkildsen for their helpful and insightful reviews of the thesis. Also, thank you Prof. Catherine Peichel for agreeing to be the Opponent.

Thank you to my co-authors and all those involved in sample and data collection for their efforts. This work would have been far from possible without these collaborations.

I am very grateful for all the members of the EvolConGen research group. Your efforts and positivity while sampling are very much appreciated. Thank you for your feedback, advice and office chatter that make days of research much brighter and fun!

Thank you to people at CIGENE for being so welcoming and helping make my visits to Ås both productive and enjoyable.

To all my friends at the University of Helsinki, thank you for making my time here more enjoyable and for your help navigating life working at the university.

Thank you Jackie, Andrew, Craig, Aina, Pierre, Camille & Antoine – your friendship has been enormously important during the last few years. Work gets done much quicker when looking forward to lunch, a walk, or evening out/in with you. Antoine, thank you for keeping the days seemingly in order during the busiest times – I am so grateful for your support. I thank my mother, father and sister for their endless love, support and encouragement during my PhD. I will be forever appreciative for the many phone chats that lifted my spirits and helped me to find motivation.

References

Aksnes, Anders, Bjarne Gjerde, and Sverre Ola Roald. 1986. “Biological, Chemical and Organoleptic Changes during Maturation of Farmed Atlantic Salmon, Salmo Salar.” *Aquaculture* 53 (1): 7–20. https://doi.org/10.1016/0044-8486(86)90295-4.

Archambeault, Sophie L., Luis R. Bärtschi, Aurélie D. Merminod, and Catherine L. Peichel. 2020. “ Adaptation via Pleiotropy and Linkage: Association Mapping Reveals a Complex Genetic Architecture within the Stickleback Eda Locus .” *Evolution Letters* 4 (4): 282–301. https://doi.org/10.1002/evl3.175.

Aykanat, T., M. Lindqvist, V. L. Pritchard, and C. R. Primmer. 2016. “From Population Genomics to Conservation and Management: A Workflow for Targeted Analysis of Markers Identified Using Genome-Wide Approaches in Atlantic Salmon Salmo Salar.” *Journal of Fish Biology*, no. 89: 2658–79. https://doi.org/10.1111/jfb.13149.

Ayllon, Fernando, Erik Kjærner-Semb, Tomasz Furmanek, Vidar Wennevik, Monica F. Solberg, Geir Dahle, Geir Lasse Taranger, et al. 2015. “The Vgll3 Locus Controls Age at Maturity in Wild and Domesticated Atlantic Salmon (Salmo Salar L.) Males.” *PLoS Genetics* 11 (11): 1–15. https://doi.org/10.1371/journal.pgen.1005628.

Ayllon, Fernando, Monica F Solberg, Kevin A Glover, Faezeh Mohammadi, Erik Kjærner-semb, Per Gunnar Fjelldal, Eva Andersson, Tom Hansen, Rolf B Edvardsen, and Anna Wargelius. 2019. “The Influence of Vgll3 Genotypes on Sea Age at Maturity Is Altered in Farmed Mowi Strain Atlantic Salmon.” *BMC Genetics* 20 (44): 1–8.

Barson, Nicola J, Tutku Aykanat, Kjetil Hindar, Matthew Baranski, Geir H Bolstad, Peder Fiske, Céleste Jacq, et al. 2015. “Sex-Dependent Dominance at a Single Locus Maintains Variation in Age at Maturity in Salmon.” *Nature* 528 (7582): 405–8. https://doi.org/https://doi.org/10.1038/nature16062.

Barton, Nicholas H., and Peter D. Keightley. 2002. “Understanding Quantitative Genetic Variation.” *Nature Reviews Genetics* 3 (1): 11–21. https://doi.org/10.1038/nrg700.

Bates, Douglas, Martin Mächler, Ben Bolker, and Steve Walker. 2015. “Fitting Linear Mixed-Effects Models Using Lme4.” *J. Stat. Softw.* 67 (1). https://doi.org/10.18637/jss.v067.i01.

Bertolotti, Alicia C, Ryan M Layer, Manu Kumar Gundappa, and Michael D Gallagher. 2020. “The Structural Variation Landscape in 492 Atlantic Salmon Genomes,” 1–36.

Bickel, Ryan D., Artyom Kopp, and Sergey V. Nuzhdin. 2011. “Composite Effects of Polymorphisms near Multiple Regulatory Elements Create a Major-Effect QTL.” *PLoS Genetics* 7 (1): 1–8. https://doi.org/10.1371/journal.pgen.1001275.

Boulding, Elizabeth G., Keng Pee Ang, J.A.K. Elliott, Frank Powell, and Lawrence R. Schaeffer. 2019. “Differences in Genetic Architecture between Continents at a Major Locus Previously Associated with Sea Age at Sexual Maturity in European Atlantic Salmon.” *Aquaculture* 500: 670–78. https://doi.org/10.1016/j.aquaculture.2018.09.025.

Bourret, Vincent, Matthew P. Kent, Craig R. Primmer, Anti Vasemägi, Sten Karlsson, Kjetil Hindar, Philip McGinnity, Eric Verspoor, Louis Bernatchez, and Sigbjørn Lien. 2013. “SNP-Array Reveals Genome-Wide Patterns of Geographical and Potential Adaptive Divergence across the Natural Range of Atlantic Salmon

(*Salmo Salar*)." *Molecular Ecology* 22 (3): 532–51. https://doi.org/10.1111/mec.12003.

Boyle, Evan A., Yang I. Li, and Jonathan K. Pritchard. 2017. "An Expanded View of Complex Traits: From Polygenic to Omnigenic." *Cell* 169 (7): 1177–86. https://doi.org/10.1016/j.cell.2017.05.038.

Bulik-Sullivan, Brendan, Po Ru Loh, Hilary K. Finucane, Stephan Ripke, Jian Yang, Nick Patterson, Mark J. Daly, et al. 2015. "LD Score Regression Distinguishes Confounding from Polygenicity in Genome-Wide Association Studies." *Nature Genetics* 47 (3): 291–95. https://doi.org/10.1038/ng.3211.

Chen, Han, Chaolong Wang, Matthew P. Conomos, Adrienne M. Stilp, Zilin Li, Tamar Sofer, Adam A. Szpiro, et al. 2016. "Control for Population Structure and Relatedness for Binary Traits in Genetic Association Studies via Logistic Mixed Models." *American Journal of Human Genetics* 98 (4): 653–66. https://doi.org/10.1016/j.ajhg.2016.02.012.

Concha, Carolina, Richard W.R. Wallbank, Joseph J. Hanly, Jennifer Fenner, Luca Livraghi, Edgardo Santiago Rivera, Daniel F. Paulo, et al. 2019. "Interplay between Developmental Flexibility and Determinism in the Evolution of Mimetic Heliconius Wing Patterns." *Current Biology* 29 (23): 3996-4009.e4. https://doi.org/10.1016/j.cub.2019.10.010.

Corre, Valérie Le, and Antoine Kremer. 2012. "The Genetic Differentiation at Quantitative Trait Loci under Local Adaptation." *Molecular Ecology* 21 (7): 1548–66. https://doi.org/10.1111/j.1365-294X.2012.05479.x.

Cousminer, Diana L., Diane J. Berry, Nicholas J. Timpson, Wei Ang, Elisabeth Thiering, Enda M. Byrne, H. Rob Taal, et al. 2013. "Genome-Wide Association and Longitudinal Analyses Reveal Genetic Loci Linking Pubertal Height Growth, Pubertal Timing and Childhood Adiposity." *Human Molecular Genetics* 22 (13): 2735–47. https://doi.org/10.1093/hmg/ddt104.

Daborn, P J, J L Yen, M R Bogwitz, G Le Goff, E Feil, S Jeffers, N Tijet, et al. 2002. "A Single P450 Allele Associated with Insecticide Resistance in <Em>Drosophila</Em>" *Science* 297 (5590): 2253 LP – 2256. https://doi.org/10.1126/science.1074170.

Day, Felix R, Deborah J Thompson, Hannes Helgason, Daniel I Chasman, Hilary Finucane, Patrick Sulem, Katherine S Ruth, et al. 2017. "Genomic Analyses Identify Hundreds of Variants Associated with Age at Menarche and Support a Role for Puberty Timing in Cancer Risk." *Nature Genetics* 49 (6). https://doi.org/10.1038/ng.3841.

Debes, Paul V, Nikolai Piavchenko, Annukka Ruokolainen, Jacqueline E Moustakas-verho, Noora Parre, Tutku Aykanat, Jaakko Erkinaro, Craig R Primmer, and Paul V Debes. 2019. "Large Single-Locus Effects for Maturation Timing Are Mediated via Body Condition in Atlantic Salmon." *BioRxiv*. https://doi.org/10.1101/780437.

Depristo, Mark A., Eric Banks, Ryan Poplin, Kiran V. Garimella, Jared R. Maguire, Christopher Hartl, Anthony A. Philippakis, et al. 2011. "A Framework for Variation Discovery and Genotyping Using Next-Generation DNA Sequencing Data." *Nature Genetics* 43 (5): 491–501. https://doi.org/10.1038/ng.806.

Erkinaro, Jaakko, Yann Czorlich, Panu Orell, Jorma Kuusela, Morten Falkegård, Maija Länsman, Henni Pulkkinen, Craig R Primmer, and Eero Niemelä. 2019. "Life History Variation across Four Decades in a Diverse Population Complex of Atlantic Salmon in a Large Subarctic River." *Canadian Journal of Fisheries and Aquatic Sciences* 76 (1): 42–55. https://doi.org/10.1139/cjfas-2017-0343.

Eu-ahsunthornwattana, Jakris, E. Nancy Miller, Michaela Fakiola, Selma M.B. Jeronimo, Jenefer M. Blackwell, and Heather J. Cordell. 2014. "Comparison of Methods to Account for Relatedness in Genome-Wide Association Studies with

Family-Based Data." *PLoS Genetics* 10 (7). https://doi.org/10.1371/journal.pgen.1004445.

Fleming, I. A., and S Einum. 2011. "Reproductive Ecology: A Tale of Two Sexes." In *Atlantic Salmon Ecology*, 35–65.

Gjedrem, Trygve, Hans Magnus Gjøen, and Bjarne Gjerde. 1991. "Genetic Origin of Norwegian Farmed Atlantic Salmon." *Aquaculture* 98 (1–3): 41–50. https://doi.org/10.1016/0044-8486(91)90369-I.

Gjerde, Bjarne. 1984. "Response to Individual Selection for Age at Sexual Maturity in Atlantic Salmon." *Aquaculture* 38 (3): 229–40. https://doi.org/10.1016/0044-8486(84)90147-9.

Gratten, J, D Beraldi, B V Lowder, A F Mcrae, P M Visscher, J M Pemberton, and J Slate. 2007. "Compelling Evidence That a Single Nucleotide Substitution in TYRP1 Is Responsible for Coat-Colour Polymorphism in a Free-Living Population of Soay Sheep," no. December 2006: 619–26. https://doi.org/10.1098/rspb.2006.3762.

Guan, Yongtao. 2014. "Detecting Structure of Haplotypes and Local Ancestry." *Genetics* 196 (3): 625–42. https://doi.org/10.1534/genetics.113.160697.

Guan, Yongtao, and Matthew Stephens. 2011. "Bayesian Variable Selection Regression for Genome-Wide Association Studies and Other Large-Scale Problems." *Annals of Applied Statistics* 5 (3): 1780–1815. https://doi.org/10.1214/11-AOAS455.

Gutierrez, Alejandro P., Krzysztof P. Lubieniecki, Steve Fukui, Ruth E. Withler, Bruce Swift, and William S. Davidson. 2014. "Detection of Quantitative Trait Loci (QTL) Related to Grilsing and Late Sexual Maturation in Atlantic Salmon (Salmo Salar)." *Marine Biotechnology* 16 (1): 103–10. https://doi.org/10.1007/s10126-013-9530-3.

Gutierrez, Alejandro P, José M Yá, Steve Fukui, Bruce Swift, and William S Davidson. 2015. "Genome-Wide Association Study (GWAS) for Growth Rate and Age at Sexual Maturation in Atlantic Salmon (Salmo Salar)." *PLoS ONE* 10 (3): 1–15. https://doi.org/10.1371/journal.pone.0119730.

Halperin, Daniel S., Calvin Pan, Aldons J. Lusis, and Peter Tontonoz. 2013. "Vestigial-like 3 Is an Inhibitor of Adipocyte Differentiation." *Journal of Lipid Research* 54 (2): 473–81. https://doi.org/10.1194/jlr.m032755.

He, Chunlin, John Holme, and Jeffrey Anthony. 2014. "SNP Genotyping: The KASP Assay BT - Crop Breeding: Methods and Protocols." In , edited by Delphine Fleury and Ryan Whitford, 75–86. New York, NY: Springer New York. https://doi.org/10.1007/978-1-4939-0446-4_7.

Hindar, Kjetil, Ian A Fleming, Philip Mcginnity, and Ola Diserud. 2006. "Genetic and Ecological Effects of Salmon Farming on Wild Salmon : Modelling from Experimental Results" 1247: 1234–47. https://doi.org/10.1016/j.icesjms.2006.04.025.

Hoekstra, H. E. 2006. "Genetics, Development and Evolution of Adaptive Pigmentation in Vertebrates." *Heredity* 97 (3): 222–34. https://doi.org/10.1038/sj.hdy.6800861.

ICES. 2011. "Report of the Workshop on Age Determination of Salmon (WKADS)."

Johnston, Susan E., Panu Orell, Victoria L. Pritchard, Matthew P. Kent, Sigbjørn Lien, Eero Niemelä, Jaakko Erkinaro, and Craig R. Primmer. 2014. "Genome-Wide SNP Analysis Reveals a Genetic Basis for Sea-Age Variation in a Wild Population of Atlantic Salmon (Salmo Salar)." *Molecular Ecology* 23 (14): 3452–68. https://doi.org/10.1111/mec.12832.

Joron, Mathieu, Riccardo Papa, Margarita Beltrán, Nicola Chamberlain, Jesús Mavárez, Simon Baxter, Moisés Abanto, et al. 2006. "A Conserved Supergene Locus Controls Colour Pattern Diversity in Heliconius Butterflies." *PLoS*

Biology 4 (10): 1831–40. https://doi.org/10.1371/journal.pbio.0040303.

Kardos, Marty, and Gordon Luikart. 2021. "The Genetic Architecture of Fitness Drives Population Viability during Rapid Environmental Change." *The American Naturalist*. https://doi.org/10.1086/713469.

Kerdaffrec, Envel, Danièle L Filiault, Arthur Korte, Eriko Sasaki, Viktoria Nizhynska, Ümit Seren, and Magnus Nordborg. 2016. "Multiple Alleles at a Single Locus Control Seed Dormancy in Swedish Arabidopsis." *ELife* 5 (3): 1–24. https://doi.org/10.7554/eLife.22502.

Kuparinen, Anna, and Jeffrey A. Hutchings. 2017. "Genetic Architecture of Age at Maturity Can Generate Either Directional or Divergent and Disruptive Harvest-Induced Evolution." *Philosophical Transactions of the Royal Society B* 372 (20160035). https://doi.org/10.1098/rstb.2016.0035.

Lamichhaney, Sangeet, Guangyi Fan, Fredrik Widemo, Ulrika Gunnarsson, Doreen Schwochow Thalmann, Marc P Hoeppner, Susanne Kerje, et al. 2015. "Structural Genomic Changes Underlie Alternative Reproductive Strategies in the Ruff (*Philomachus Pugnax*)." *Nature Genetics* 48 (1): 84–88. https://doi.org/10.1038/ng.3430.

Lamichhaney, Sangeet, Angela P. Fuentes-Pardo, Nima Rafati, Nils Ryman, Gregory R. McCracken, Christina Bourne, Rabindra Singh, Daniel E. Ruzzante, and Leif Andersson. 2017. "Parallel Adaptive Evolution of Geographically Distant Herring Populations on Both Sides of the North Atlantic Ocean." *Proceedings of the National Academy of Sciences of the United States of America* 114 (17): E3452–61. https://doi.org/10.1073/pnas.1617728114.

Legarra, Andres, Anne Ricard, and Luis Varona. 2018. "GWAS by GBLUP: Single and Multimarker EMMAX and Bayes Factors, with an Example in Detection of a Major Gene for Horse Gait." *G3 (Bethesda, Md.)* 8 (7): 2301–8. https://doi.org/10.1534/g3.118.200336.

Linnen, Catherine R, Yu-Ping Poh, Brant K Peterson, Rowan D H Barrett, Joanna G Larson, Jeffrey D Jensen, and Hopi E Hoekstra. 2013. "Adaptive Evolution of Multiple Traits through Multiple Mutations at a Single Gene." *Science* 339 (6125): 1312–16. https://doi.org/10.1126/science.1233213.

Listgarten, Jennifer, Christoph Lippert, Carl M Kadie, Robert I Davidson, Eleazar Eskin, and David Heckerman. 2012. "Improved Linear Mixed Models for Genome-Wide Association Studies." *Nature Methods* 9 (May): 525. https://doi.org/http://10.0.4.14/nmeth.2037.

Liu, Xuanyao, Yang I. Li, and Jonathan K. Pritchard. 2019. "Trans Effects on Gene Expression Can Drive Omnigenic Inheritance." *Cell* 177 (4): 1022-1034.e6. https://doi.org/10.1016/j.cell.2019.04.014.

Livraghi, Luca, Arnaud Martin, Melanie Gibbs, Nora Braak, Saad Arif, and Casper J. Breuker. 2018. *CRISPR/Cas9 as the Key to Unlocking the Secrets of Butterfly Wing Pattern Development and Its Evolution*. *Advances in Insect Physiology*. 1st ed. Vol. 54. Elsevier Ltd. https://doi.org/10.1016/bs.aiip.2017.11.001.

Loh, Po Ru, Gaurav Bhatia, Alexander Gusev, Hilary K. Finucane, Brendan K. Bulik-Sullivan, Samuela J. Pollack, Teresa R. De Candia, et al. 2015. "Contrasting Genetic Architectures of Schizophrenia and Other Complex Diseases Using Fast Variance-Components Analysis." *Nature Genetics* 47 (12): 1385–92. https://doi.org/10.1038/ng.3431.

Loh, Po Ru, George Tucker, Brendan K. Bulik-Sullivan, Bjarni J. Vilhjálmsson, Hilary K. Finucane, Rany M. Salem, Daniel I. Chasman, et al. 2015. "Efficient Bayesian Mixed-Model Analysis Increases Association Power in Large Cohorts." *Nature Genetics* 47 (3): 284–90. https://doi.org/10.1038/ng.3190.

Luconi, Michaela, Vinicio Carloni, Fabio Marra, Pietro Ferruzzi, Gianni Forti, and Elisabetta Baldi. 2004. "Increased Phosphorylation of AKAP by Inhibition of

Phosphatidylinositol 3-Kinase Enhances Human Sperm Motility through Tail Recruitment of Protein Kinase A." *Journal of Cell Science* 117 (Pt 7): 1235–46. https://doi.org/10.1242/jcs.00931.

McClure, Carol A., K. Larry Hammell, Mark Moore, Ian R. Dohoo, and Holly Burnley. 2007. "Risk Factors for Early Sexual Maturation in Atlantic Salmon in Seawater Farms in New Brunswick and Nova Scotia, Canada." *Aquaculture* 272 (1–4): 370–79. https://doi.org/10.1016/j.aquaculture.2007.08.039.

Miki, Kiyoshi, William D Willis, Paula R Brown, Eugenia H Goulding, Kerry D Fulcher, and Edward M Eddy. 2002. "Targeted Disruption of the Akap4 Gene Causes Defects in Sperm Flagellum and Motility." *Developmental Biology* 248 (2): 331–42. https://doi.org/10.1006/dbio.2002.0728.

Mobley, K.B., Hanna Granroth-Wilding, Mikko Ellmén, Panu Orell, Jaakko Erkinaro, and Craig R. Primmer. 2020. "Time Spent in Distinct Life History Stages Has Sex-Specific Effects on Reproductive Fitness in Wild Atlantic Salmon." *Molecular Ecology* 29 (6): 1173–84. https://doi.org/10.1111/mec.15390.

Mobley, Kenyon B., Tutku Aykanat, Yann Czorlich, Andrew House, Johanna Kurko, Antti Miettinen, Jacqueline Moustakas-Verho, et al. 2021. *Maturation in Atlantic Salmon (Salmo Salar, Salmonidae): A Synthesis of Ecological, Genetic, and Molecular Processes*. *Reviews in Fish Biology and Fisheries*. Vol. 0123456789. https://doi.org/10.1007/s11160-021-09656-w.

Mohamed, Amin R, Klara L Verbyla, Hawlader A Al-mamun, Sean Mcwilliam, Bradley Evans, Harry King, Peter Kube, and James W Kijas. 2019. "Polygenic and Sex Specific Architecture for Two Maturation Traits in Farmed Atlantic Salmon." *BMC Genomics* 20 (139): 1–12.

Moore, Jean Sébastien, Vincent Bourret, Mélanie Dionne, Ian Bradbury, Patrick O'Reilly, Matthew Kent, Gérald Chaput, and Louis Bernatchez. 2014. "Conservation Genomics of Anadromous Atlantic Salmon across Its North American Range: Outlier Loci Identify the Same Patterns of Population Structure as Neutral Loci." *Molecular Ecology* 23: 5680–97. https://doi.org/10.1111/mec.12972.

Oomen, Rebekah A, Anna Kuparinen, and Jeffrey A Hutchings. 2020. "Consequences of Single-Locus and Tightly Linked Genomic Architectures for Evolutionary Responses to Environmental Change." *Journal of Heredity*, 319–32. https://doi.org/10.1093/jhered/esaa020.

Perry, John R B, Felix Day, Cathy E Elks, Patrick Sulem, Deborah J Thompson, Teresa Ferreira, Chunyan He, et al. 2014. "Parent-of-Origin-Specific Allelic Associations among 106 Genomic Loci for Age at Menarche." *Nature* 514 (July): 92. https://doi.org/http://10.0.4.14/nature13545.

Price, Alkes L, Noah A Zaitlen, David Reich, and Nick Patterson. 2010. "New Approaches to Population Stratification in Genome-Wide Association Studies." *Nature Reviews Genetics* 11 (June): 459. https://doi.org/http://10.0.4.14/nrg2813.

Pritchard, Jonathan K., and Anna Di Rienzo. 2010. "Adaptation - Not by Sweeps Alone." *Nature Reviews Genetics* 11 (10): 665–67. https://doi.org/10.1038/nrg2880.

Pritchard, V. L., Makinen H., Vähä J.-P., Erkinaro J., Orell P., and C.R. Primmer. 2018. "Genomic Signatures of Fine-Scale Selection in Atlantic Salmon Suggest Involvment of Sexual Maturation, Energy Homeostatis, and Immune Defence-Related Genes." *Molecular Ecology* 27: 2560–75. https://doi.org/10.1111/mec.14705.

Purcell, Shaun M., Naomi R. Wray, Jennifer L. Stone, Peter M. Visscher, Michael C. O'Donovan, Patrick F. Sullivan, Douglas M. Ruderfer, et al. 2009. "Common Polygenic Variation Contributes to Risk of Schizophrenia and Bipolar Disorder." *Nature* 460 (7256): 748–52. https://doi.org/10.1038/nature08185.

Reed, T. E., P. Prodöhl, R. Hynes, T. Cross, A. Ferguson, and P. McGinnity. 2015. "Quantifying Heritable Variation in Fitness-Related Traits of Wild, Farmed and Hybrid Atlantic Salmon Families in a Wild River Environment." *Heredity* 115 (2): 173–84. https://doi.org/10.1038/hdy.2015.29.

Reinton, Nils, Philippe Collas, Trine B Haugen, Bjørn S Skålhegg, Vidar Hansson, Tore Jahnsen, and Kjetil Taskén. 2000. "Localization of a Novel Human A-Kinase-Anchoring Protein, HAKAP220, during Spermatogenesis." *Developmental Biology* 223 (1): 194–204. https://doi.org/https://doi.org/10.1006/dbio.2000.9725.

Rodríguez-Leal, Daniel, Zachary H Lemmon, Jarrett Man, Madelaine E Bartlett, and Zachary B Lippman. 2017. "Engineering Quantitative Trait Variation for Crop Improvement by Genome Editing." *Cell* 171 (2): 470-480.e8. https://doi.org/10.1016/j.cell.2017.08.030.

Sander, Jeffry D, and J Keith Joung. 2014. "CRISPR-Cas Systems for Editing, Regulating and Targeting Genomes." *Nature Biotechnology* 32 (4): 347–55. https://doi.org/10.1038/nbt.2842.

Sedeek, Khalid E M, Ahmed Mahas, and Magdy Mahfouz. 2019. "Plant Genome Engineering for Targeted Improvement of Crop Traits." *Frontiers in Plant Science* 10 (February): 114. https://doi.org/10.3389/fpls.2019.00114.

Sham, Pak C., and Shaun M. Purcell. 2014. "Statistical Power and Significance Testing in Large-Scale Genetic Studies." *Nature Reviews Genetics* 15 (5): 335–46. https://doi.org/10.1038/nrg3706.

Taranger, Geir Lasse, Manuel Carrillo, Rüdiger W. Schulz, Pascal Fontaine, Silvia Zanuy, Alicia Felip, Finn Arne Weltzien, et al. 2010. "Control of Puberty in Farmed Fish." *General and Comparative Endocrinology* 165 (3): 483–515. https://doi.org/10.1016/j.ygcen.2009.05.004.

The International Mouse Knockout Consortium. 2007. "A Mouse for All Reasons." *Cell* 128 (1): 9–13. https://doi.org/10.1016/j.cell.2006.12.018.

Tishkoff, Sarah A., Floyd A. Reed, Alessia Ranciaro, Benjamin F. Voight, Courtney C. Babbitt, Jesse S. Silverman, Kweli Powell, et al. 2007. "Convergent Adaptation of Human Lactase Persistence in Africa and Europe." *Nature Genetics* 39 (1): 31–40. https://doi.org/10.1038/ng1946.

Turelli, Michael, and N. H. Barton. 2004. "Polygenic Variation Maintained by Balancing Selection: Pleiotropy, Sex-Dependent Allelic Effects and G × E Interactions." *Genetics* 166 (2): 1053–79. https://doi.org/10.1534/genetics.166.2.1053.

Varshney, Gaurav K, Jing Lu, Derek E Gildea, Haigen Huang, Wuhong Pei, Zhongan Yang, Sunny C Huang, et al. 2013. "A Large-Scale Zebrafish Gene Knockout Resource for the Genome-Wide Study of Gene Function." *Genome Research* 23 (4): 727–35. https://doi.org/10.1101/gr.151464.112.

Waples, Ryan K, Wesley A Larson, and Robin S Waples. 2016. "Estimating Contemporary Effective Population Size in Non-Model Species Using Linkage Disequilibrium across Thousands of Loci." *Heredity* 117: 233–440. https://doi.org/10.1038/hdy.2016.60.

Woronik, Alyssa, Kalle Tunström, Michael W. Perry, Ramprasad Neethiraj, Constanti Stefanescu, Maria de la Paz Celorio-Mancera, Oskar Brattström, et al. 2019. "A Transposable Element Insertion Is Associated with an Alternative Life History Strategy." *Nature Communications* 10 (1). https://doi.org/10.1038/s41467-019-13596-2.

Wucherpfennig, Julia I., Craig T. Miller, and David M. Kingsley. 2019. "Efficient CRISPR-Cas9 Editing of Major Evolutionary Loci in Sticklebacks." *Evolutionary Ecology Research* 20 (1–3): 1–29.

Xu, Hanli, and Yongtao Guan. 2014. "Detecting Local Haplotype Sharing and

Haplotype Association." *Genetics* 197 (3): 823–38. https://doi.org/10.1534/genetics.114.164814.

Yang, Jian, Teresa Ferreira, Andrew P Morris, Sarah E Medland, Pamela A F Madden, Andrew C Heath, Nicholas G Martin, et al. 2012. "Conditional and Joint Multiple-SNP Analysis of GWAS Summary Statistics Identifies Additional Variants Influencing Complex Traits." *Nature Genetics* 44 (4): 369–75. https://doi.org/10.1038/ng.2213.

Yang, Jian, S. Hong Lee, Michael E. Goddard, and Peter M. Visscher. 2011. "GCTA: A Tool for Genome-Wide Complex Trait Analysis." *American Journal of Human Genetics* 88 (1): 76–82. https://doi.org/10.1016/j.ajhg.2010.11.011.

Yano, Ayaka, Barbara Nicol, Elodie Jouanno, Edwige Quillet, and Alexis Fostier. 2012. "The Sexually Dimorphic on the Y-Chromosome Gene (SdY) Is a Conserved Male-Specific Y-Chromosome Sequence in Many Salmonids." *Evolutionary Applications*. https://doi.org/10.1111/eva.12032.

Yeaman, Sam. 2013. "Genomic Rearrangements and the Evolution of Clusters of Locally Adaptive Loci." *Proceedings of the National Academy of Sciences of the United States of America* 110: E1743-51. https://doi.org/10.1073/pnas.1219381110.

Yeaman, Sam, Simon Aeschbacher, and Reinhard Bürger. 2016. "The Evolution of Genomic Islands by Increased Establishment Probability of Linked Alleles." *Molecular Ecology* 25 (11): 2542–58. https://doi.org/https://doi.org/10.1111/mec.13611.

Yeaman, Sam, and Michael C. Whitlock. 2011. "The Genetic Architecture of Adaptation under Migration-Selection Balance." *Evolution* 65 (7): 1897–1911. https://doi.org/10.1111/j.1558-5646.2011.01269.x.

Zhou, Wei, Jonas B. Nielsen, Lars G. Fritsche, Rounak Dey, Maiken E. Gabrielsen, Brooke N. Wolford, Jonathon LeFaive, et al. 2018. "Efficiently Controlling for Case-Control Imbalance and Sample Relatedness in Large-Scale Genetic Association Studies." *Nature Genetics* 50 (9): 1335–41. https://doi.org/10.1038/s41588-018-0184-y.